QA
76
M94
S1-Ref.

M U D P I E no. 13

Museum and University Data, Program and Information Exchange

[Editor's note--Harriet R. Meadow, of the Flora North America Staff, Department of Botany, National Museum of Natural History, Washington, D. C. 20560, has been preparing surveys of various information systems for the use of the FNA Secretariat. While these summaries are informal, and were not written for general distribution, I think they are of considerable interest, and those of the series I consider of particular significance to MUDPIE readers will be included in this and following issues. --JAP].

INFORMATION SYSTEMS SURVEY
"GIPSY"
[General Information Processing System]

The GIPSY system is a generalized information processing package which provides facility for creating, maintaining and querying fielded files of data. A high level user language is provided. It was written in assembly language for the IBM 360 and operates under full OS in either batch or teleprocessing mode. The 1050 or 2741 may be used as the teleprocessing terminal.

The system was built at the University of Oklahoma under Dr. James Sweeney with partial support from the federal government. It is currently operational, is available at no cost to selected users, and is generally available as a rental product.

The system does not provide for use of data element value dictionaries either as input validation devices or as query expansion devices.

FILE ORGANIZATION -- Data files are stored on disk in a sequential mode. Records are variable length; each may be composed of a variable number of variable length fields. A field consists of a tag or identifier followed by a data value or a text expression. No hierarchical organization of data fields is provided. No repeating fields are allowed.

The content of each data file is described to the system so that users of the system need be concerned only with the logical organization of their data and may refer to data items by their names. They need not be aware of physical storage of the data within the computer.

FILE CREATION -- GIPSY uses a utility program to create a new data file from card images, based upon the file description previously stored in the system. During this process syntactic error analysis is performed but no error analysis of content is provided.

SMITHSONIAN
JAN 11 1971
LIBRARIES

FILE MAINTENANCE -- An existing GIPSY file may be updated by adding new records, deleting records or changing the content of existing records. Changes may include adding a field, adding text to the end of an existing field, replacing an existing field or deleting a field.

RETRIEVAL -- The user requests a search using a high level language to specify a Boolean logic statement made up of elements of the form: data field - relational operator - data value. Text fields may be scanned for a match in any position and for any predefined number of occurrences of the data value. They can be searched on alphabetic limits (SAM through STE) and for pairs of words within a fixed distance of each other. Numeric fields can be searched on value (=, , , range) and two numeric fields in the same record can be compared.

Search is performed against a single file in the sequential mode. Searches may be performed iteratively with an initial general search identifying a relevant subset of the total file and subsequent more specific searches being performed against those records contained in the subset.

Listings of complete data records or of selected data items may be requested from GIPSY following a search. It is possible to compute sums, counts and averages and to output maximum and minimum values for a field. Date may be sorted before it is printed. The output phase may also be used to create an output file consisting of fixed length, fixed fielded records for use in another system.--Harriet R. Meadow.

FIRE AT SMITHSONIAN DESTROYS COMPUTER EXHIBIT

[From Washington Post, Oct. 1, 1970]

An early morning blaze on the third floor of the Smithsonian Institution's Museum of History and Technology destroyed a computer exhibit yesterday.

Damage was estimated by a fire marshal at "several hundred thousand dollars." Smithsonian officials later said they could not determine losses until they assessed the extent of structural damage caused by heat and water and the cost of restoring collector items damaged but not ruined in the fire. There were no injuries.

By the time the blaze was extinguished, a computer terminal valued at $50,000 had been destroyed. Also lost in the fire were furnishings in the computer exhibit, at which visitors--mostly children--took turns operating a computer keyboard to answer questions and compete in a mechanical game of tic-tac-toe.

Fire marshals did not pinpoint the cause of the fire, but they indicated it may have resulted from a short circuit in the computer terminal. The main computer is located elsewhere, they said.

AVAILABLE PROGRAMS-FORTRAN IV

The following statistical programs are presently available in the disk library of the Northern Illinois University Computer Services. Program write-ups and/or Fortran IV listings are available for distribution without charge. Contact: Dr. Jerrold H. Zar, Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115.

NIU Number	Name	Title
RZ001	CON1	Conversion of Data
RZ002	REP1	Converting BCD (26KP) Deck to EBCDIC (29KP) Deck
RZ003	REP2	Reproduce with Numbering by 10
RZ004	REP3	Listing and Reproducing
RZ005	SM1	Sample Mean and Measures of Dispersion
RZ006	SM2	Sample Mean, Median, and Measures of Dispersion
RZ007	SM3	Sample Mean and Measures of Dispersion for each Change in ID
RZ008	INF1	Information Content of N Individuals in S Categories
RZ009	CHS1	Chi-Square Contingency Table
RZ010	CHS2	Chi-Square Contingency tables
RZ011	FET1	Fisher Exact Test for 2 X N Tables
RZ012	MWU1	Mann-Whitney U Test
RZ013	FAV1	Two-Way Analysis of Variance by Ranks (Friedman's Test)
RZ014	KSD1	Kolmogorov-Smirnov Two-Sample Test
RZ015	KSD2	Kolmogorov-Smirnov Two-Sample Tests
RZ016	SRC1	Spearman Rank Correlation
RZ017	KRC1	Kendall Rank Correlation
RZ018	KCC1	Kendall Coefficient of Concordance
RZ019	LRC1	Simple Linear Regression and Correlation (includes optional logarithmic transformations)
RZ020	MRC1	Multiple Regression and Correlation
RZ021	MRP1	Predicting Y Values from Multiple Regression Equations
RZ022	POL1	Polynomial Curve Fitting and Analysis
RZ023	NLR1	Nonlinear Regression of the Model Y+=B(1)+B(2)*X(1)**B(3)+SIGMA(B(J)*X(J-2))
RZ024	NLP1	Predicting Y Values from Nonlinear Regressions of the Model of NLR1
RZ025	CIR1	Statistics of Circular Distributions
RZ026	CIR2	Two-sample Tests for Circular Distributions

COMPUTER-BASED INFORMATION NETWORKS

Volume 4 of the Annual Review of Information Science and Technology has been published by the American Society for Information Science (Britannica, Chicago, 1969, pp. 1-547). It includes an article on "Information Networks" by Carl F. J. Overhage (pp. 339-378) that summarizes computer network exchanges up to and including 1968. The idea that a network connecting museums and universities to facilitate exchange of data, programs and information would be useful has always been the motive force in MUDPIE, so this article is particularly pertinent. After reading this article I think perhaps we are closer to achieving such a network than I had thought.

Overhage cites one reference that estimates the number of computing systems providing access from remote terminals at over 1000. He mentions examples of various kinds of computer-communications systems, saying "it is the combination of computers, data circuits, and user terminals that constitutes an information network in the modern sense." Some of the existing or developing networks are: a network of 35 interconnected computers at 16 locations organized by the Advanced Research Projects Agency of the Department of Defense, and "controlled by research groups that have agreed to accept a single network protocol in order to participate in network experiments;" a regional network in western Pennsylvania and West Virginia for 10 colleges and universities, established by Carnegie-Mellon University; a network "designed to stimulate computer use in research" involving the University of Iowa and 10 area colleges; the Triangle Universities Computer Center, which serves Duke, North Carolina, North Carolina State and about 35 smaller colleges; three different networks involving SUNY in various combinations with other New York institutions; a network centered at the University of Georgia, which, among other things, "takes data on this week's football-playing performance of next week's opponents and produces a report on [their] strengths and weaknesses...;" and so on. Overhage refers to a network directory being assembled by EDUCOM that will list over 50 university based information networks. Perhaps the next step for MUDPIE is to see what is involved in getting representatives together with the hopes of beginning interchange between a few of these already established networks.--JAP

COMPUTERS REVIEWED

The October, 1970, issue of Science Journal (London) is almost entirely devoted to computers and computing, and provides a lot of interesting reading. MUDPIE readers will find the following of specific value:

Living with Computers [pp. 35-38]. By A. S. Douglas.

The Intelligent Machine [pp. 50-54]. By Donald Michie.

Computer Networks [pp. 60-64]. By D. L. A. Barber.

Computers in Medicine [pp. 81-86]. By Thomas Tyler.

Computers in Education [pp. 89-93]. By E. B. James.

RECENT LITERATURE

Bolt, A. B. & M. E. Wardle. Communicating with a Computer. Cambridge Univ. Press, 1970, 80 pp. [A grade school-high school level introduction to computers; quite readable and informative for an initial plunge into computerization.]

Collette, B. B. & N. V. Parin. Needlefishes (Belonidae) of the Eastern Atlantic Ocean. Atlantide Report no. 11, 1970, pp. 1-60. [Noteworthy for graphs, all of which were plotted by time-shared computer on teletype.]

Hull, T. E. & D. D. F. Day. Computers and problem solving. Addison-Wesley, 1970, pp. xii + 276. [This is an introduction to FORTRAN programming, plus lots of interesting little problems and sidelights.]

Johnson, L. A. S. Rainbow's end: the quest for an optimal taxonomy. Systematic Zoology, 19, 1970, pp. 203-239.

Sokal, R. R. & F. J. Rohlf. The intelligent ignoramus, an experiment in numerical taxonomy. Taxon, 19, 1970, 305-319. [In an attempt to demonstrate that anyone can do taxonomic work successfully the authors leave themselves wide open to serious doubts as to whether they are eligible for inclusion in the group.]

Stark, P. A. Introduction to Numerical Methods. MacMillan, London, 1970, pp. xii + 334. [This is of value only to biologists who are also competent mathematicians.]

Wilkinson, C. Adding a point to a principal coordinates analysis. Systematic Zoology, 19, 1970, 258-263.

December, 1970
Division of Reptiles and Amphibians
National Museum of Natural History
Washington, D. C. 20560